CS3810 Fall 2017
Homework 3
Due: Thursday, Oct 5, 2017

This assignment will work with extensive data files from the US Dept of Transportation, Bureau of Transportation Statistics.

Data for this assignment can be found in a variety of subpages that begin at:
`https://www.transtats.bts.gov/`

Before doing any work with the database, determine tables based on the contents of the files you need, identify the relationships among the tables, determine the primary and foreign keys. In other words, you're going to be working backwards the Data Library: Aviation. The data you need to download start at this location:

`https://www.transtats.bts.gov/DL_SelectFields.asp`
Air Carriers : T-100 Domestic Market (U.S. Carriers)

- Colorado, 2017, June
- Select all fields
- Documentaion
- Term

This next collection has details on all the airlines, world-wide. There is duplication with the T-100 file. Normalization will take care of this.

Aviation Support Tables : Carrier Decode

- Select all fields
- Documentaion
- Term

In each download there are 3 files - Terms, Documentation and the data file. All three are in CSV (comma-delimited, plain text) that can be read by Excel.

Unzip the files, read the documentation, understand what the terms are.

Now put the data file into Excel (call this the original table) - much easier to read and analyze the data.

The first row (header row) of the data file are the names of the columns in the table and will be the names of the columns in your tables. But, this file is a flat file, and you need to put the data into 3rd Normal form before you start creating the tables. As you go thru the normalization process, you'll need to create additional columns. Name these new columns something meaningful.

After you've completed the normalization, create a sheet for each of the tables you defined and copy the columns from the original table into the appropriate sheet.

For example, in the data file, the column UNIQUE_CARRIER appears more than 10 times. Hence in the normalization process, you'll need to create a separate table that contains UNIQUE_CARRIER, UNIQUE_CARRIER_ENTITY and UNIQUE_CARRIER_NAME (please, don't use all caps!)

When finished, save each sheet in CSV format (plain text) with the name of the table the data will be imported into.

Trust me, you do not want to load each row by hand. Download the file(s) and format for your tables. Then do a bulk import (or bulk load) into the database. This is also known as populating a database. You can find more details and examples here:

`https://www.postgresql.org/docs/9.4/static/populate.html`

CSV files can imported directly into PostGres **after** you've defined and created the tables. Read this for instructions:
`http://www.postgresqltutorial.com/import-csv-file-into-posgresql-table/`

Do This:
Create three plain text files of SQL:

- The CREATE statements that create the tables in a file named `HW3Create.sql`
- The COPY statements that load the tables in a file named `HW3Load.sql`

To Turn In:
Upload to your github account the following:

- The Excel workbook that contains the results of the normalization analysis.
- HW3Create.sql
- HW3Load.sql

What's missing? Queries! That's the next assignment, which depends on this one.